

Rochester Institute of Technology

RIT Scholar Works

Articles

Faculty & Staff Scholarship

11-18-2020

Brightness matching in optical see-through augmented reality

Michael Murdoch

Rochester Institute of Technology

Follow this and additional works at: <https://scholarworks.rit.edu/article>

Recommended Citation

Michael J. Murdoch, "Brightness matching in optical see-through augmented reality," J. Opt. Soc. Am. A 37, 1927-1936 (2020) <https://doi.org/10.1364/JOSAA.398931>

This Article is brought to you for free and open access by the Faculty & Staff Scholarship at RIT Scholar Works. It has been accepted for inclusion in Articles by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

Brightness matching in optical see-through augmented reality

MICHAEL J. MURDOCH 

Munsell Color Science Laboratory, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, New York 14623, USA (e-mail: michael.murdoch@mail.rit.edu)

Received 29 May 2020; revised 14 October 2020; accepted 15 October 2020; posted 15 October 2020 (Doc. ID 398931); published 18 November 2020

A visual experiment using a beam-splitter-based optical see-through augmented reality (OST-AR) setup tested the effect of the size and alignment of AR overlays with a brightness-matching task using physical cubes. Results indicate that more luminance is required when AR overlays are oversized with respect to the cubes, showing that observers discount the AR overlay to a greater extent when it is more obviously a transparent layer. This is not explained by conventional color appearance modeling but supports an AR-specific model based on foreground-background discounting. The findings and model will help determine parameters for creating convincing AR manipulation of real-world objects. © 2020 Optical Society of America

<https://doi.org/10.1364/JOSAA.398931>

Provided under the terms of the [OSA Open Access Publishing Agreement](#)

1. INTRODUCTION

Optical see-through (OST) augmented reality (AR) is an example of extended reality and mixed reality in which virtual content, such as computer graphics, text, or video, is presented to a user via a transparent optical system through which the real world remains visible. A defining characteristic of OST-AR is the optical blending of the virtual, transparent AR overlay and the real-world scene visible through the system. The AR overlay may be anchored in real-world coordinates with robust sensing of user and object position, which allows two use cases: inserting virtual objects into the real world and manipulation of real-world objects with aligned AR overlays. The latter case is the main focus of the present investigation, which addresses the visual perception of object brightness as it is manipulated by AR overlays of varying luminance, size, and alignment. Applications for AR manipulation of real-world objects include medicine, education, retail, and entertainment, all of which will benefit from a clearer understanding of how users perceive and interpret transparent, luminous AR overlays in their field of view.

In previous research, the physical blending of AR foreground overlays and real-world backgrounds was analyzed with a commercial AR system, showing the distorting shifts in chromaticity caused by backgrounds bleeding through the transparent display [1]. For spatial AR (aka projection mapping), a colorimetric compensation scheme was proposed to adjust the projected AR content to account for the distortion of the background object [2]. These studies have addressed the physics of blending but not of color appearance effects. Taking the next step, Hassani and Murdoch performed color-matching experiments between AR overlays with different physical backgrounds [3,4]. They found

limited success modeling the results using CIECAM02, a color appearance model that incorporates surround and luminance level effects, but that does not account for transparency [5]. It appears that AR users are able to interpret the AR as a transparent overlay, in which case they can partially discount the effect of the background physically bleeding through the transparent display. Hassani proposed a model structure that is discussed below [4].

Hassani's studies asked observers to focus on the AR foreground overlays and implicitly discount the background, but it is expected that observers can also do the opposite: focus on the background and discount the AR foreground. This is a ubiquitous experience as people ignore reflections on glass while blithely interpreting the scene behind. Sixty years ago, Fry and Alpern developed the concept of veiling luminance [6], defined clearly by Gilchrist and Jacobsen as "a homogeneous sheet of light that is used to 'veil,' or cover, some other pattern" [7]. Generally speaking, if an overlay (created in some experiments with a beam splitter overlaying a uniform field of light on a scene) is uniform and not aligned with scene elements, it can be discounted as veiling luminance, meaning observers have no trouble assessing object lightness and other characteristics through the veil; another way to say this is that lightness constancy is preserved in such arrangements. Fry and Alpern proposed that an overlay extending at least 0.25 degrees of visual angle beyond the edges of the object is likely to be perceived as a veil; however, a related study showed that with textureless stimuli, as the extension of a transparent filter was reduced, perceived lightness was enhanced, similarly to veiling luminance [8].

OST-AR displays are physically transparent, but it is well known that the percept of transparency can be evoked by certain spatial luminance arrangements, regardless of whether they involve transparent filters or simulations. Kingdom provides two excellent reviews of research on transparency perception, emphasizing in the earlier review the important role of the physical correctness, or at least plausibility, of adjacent luminance ratios and luminance edges [9,10]. In OST-AR displays, the overlay is by its nature physically transparent, but it seems that the perception of transparent veiling luminance can yield to something like the Gelb effect (in which an illumination difference is perceptually attributed to an object's surface reflectance [10]) if the AR overlay is well aligned with background object edges. Additionally, there is a link between transparency perception and simultaneous contrast in the case of a transparent overlay smaller than a background object [11], which may be broken in cases where there is a mismatch in texture between foreground and background [12].

Many researchers have studied the perception of transparent filters atop various background color arrangements, using episcotisters, subtractive filters, and with physically plausible simulations. Metelli showed that when two different background grays are visible, with a filter partially over both, the transparency can be estimated visually and modeled algebraically [13]. D'Zmura et al. showed a similar effect with isoluminant colored filter and background combinations [14]. In both cases, the resulting proximal color can be modeled as a transparency-weighted blend of foreground and background values as in the following conceptual formula with τ representing transparency:

$$\text{Blend} = (1 - \tau) \text{Foreground} + \tau \text{Background}, 0 \leq \tau \leq 1. \quad (1)$$

For transparency, the weights of foreground and background typically sum to unity. However, because OST-AR systems add light via a transparent display, and the transmission of the background through the display is constant regardless of AR foreground content, the blend would be expected to comprise unit amounts of both AR foreground and transmitted background. However, Hassani found that a non-physical weighted sum of the CIE tristimulus values of foreground and background was required to model visual matching results, and that the weighting depends on the relative luminance and complexity of the foreground stimuli [4]. The structure of her formulation is as follows, where α and β are weighting scalars for the XYZ tristimulus values of the foreground (FG) and background (BG), respectively, summing to an effective XYZ corresponding to a visual match:

$$XYZ_{\text{eff}} = \alpha XYZ_{\text{FG}} + \beta XYZ_{\text{BG}}. \quad (2)$$

In AR experiments matching foreground colors, $\alpha : \beta$ ratios of between 2:1 and 3:1 were found, with β smaller than α because observers were instructed to match the color of foreground AR overlays, implicitly discounting the background. Without discounting, one would expect a physical mix, $\alpha = \beta = 1$. In a hypothetical case of fully discounting the background, $\beta = 0$.

Ideally, an AR system would provide perfect alignment for virtual objects and overlays with respect to real-world objects. However, practicalities suggest that this will remain difficult

for arbitrary real-world objects and complex virtual images in real-time graphics systems. Thus, the effect of alignment and size of AR overlays on perceived color remains important to understand. This is specifically important if spatial differences can make the difference between users interpreting AR overlays as object colors versus veiling luminance.

A. Experiment Development

Based on this background, an experiment was devised to study the effect of AR overlay alignment and size on perceived transparency and veiling luminance. Contrasting with previous studies, attention turns from the matches with the AR foreground to matches with a physical, 3D background. It is expected that the alignment and size of an AR overlay affect how it is perceived and discounted. Two related hypotheses arise. First, discounting will be minimal for a tightly fit AR overlay that is more likely to be interpreted as integral to the background object. Second, discounting will be stronger if an AR overlay is oversized enough to be interpreted as veiling luminance, meaning a brighter overlay will be required to make a visual match. Further, it is expected that if Hassani's model structure [Eq. (2)] applies, the $\alpha : \beta$ ratio will be inverted (meaning $\alpha < \beta$) as observers discount the foreground to match the background objects.

The general objective is to study AR overlays and backgrounds of arbitrary color; however, for the sake of a tractable starting point, luminance was chosen as a single-dimension starting point. The physical characteristic of luminance is related to percepts of brightness and lightness, terms that carry differing connotations in various domains of the study of perception as they are used to explain visual interrelationships between object properties and illumination [15]. A review by Kingdom, already mentioned, summarizes some of the many interpretations of these terms, pointing out that in the absence of illumination differences between pairs of stimuli, brightness and lightness are synonymous [10]. Herein, the definitions follow conventions used in color science [16], with quotations from the CIE International Lighting Vocabulary [17]: brightness is the "attribute of a visual perception according to which an area appears to emit, or reflect, more or less light." Lightness is defined for a related color, meaning one viewed in context, as the "brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting."

While the nuances of brightness and lightness can be ambiguous to both experts and novices, specific knowledge of these definitions was not essential to the experimental task. A simple brightness-matching task was chosen with the intention that brightness could be matched visually, by adjusting luminance, without the details of these definitions. A matching task with one dimension of control is simple for observers to execute, and while lightness matching might seem appropriate to some readers, the ambiguities of reflective surfaces and transparent overlays would have made this complicated to explain. Gilchrist, while presenting brightness as an impossible to measure proximal quantity, points out that in an unambiguous situation (i.e., without an illumination difference) matches of brightness and lightness are equivalent [18]. Thus, in the present experiment

the distinction between these tasks may be minimal, and observers were asked to match brightness; specific instructions used for the experiment are quoted in Section 2.C.

In a pilot study, an initial investigation was conducted using transparent AR overlays to manipulate the apparent brightness of real-world cubes, each painted a different shade of gray [19]. Brightness-matching results partially supported the aforementioned hypotheses, but there was a difference in results between cubes with different physical reflectance that was confounded with presentation order, meaning it was unclear whether the difference was due to reflectance or the result of learning effects. In order to clarify the pilot results and get a more precise understanding of the perception of real-world 3D objects with AR overlays, a new brightness-matching experiment was performed. The present experiment uses the same painted cubes with a new interface and improved methodology, including counterbalanced presentation order, observer training, and repeated presentations.

2. METHODS

A. Apparatus

The experiment was conducted using a desktop AR setup, essentially an example of the nineteenth-century Pepper's Ghost illusion, previously described in [3]. Depicted in Fig. 1, the AR setup comprises a viewing booth with controllable lighting, a beam splitter (half-silvered mirror), and an LCD display. The observer sees an optical blend of illuminated real-world objects in the viewing booth and virtual content displayed on the LCD, with the virtual content appearing transparent and emissive in front of the real-world objects. Viewed monocularly to eliminate stereo cues indicating that it is flat, virtual content that is well aligned with the 3D real-world objects from the observer's eye position appears to merge with the objects themselves, changing their appearance. To ensure this fixed eye position, the AR setup was outfitted with a viewing cone with a 2 cm hole and a rubber eyecup from a Nikon SLR camera. Observers performed the experiment monocularly from this constrained position.

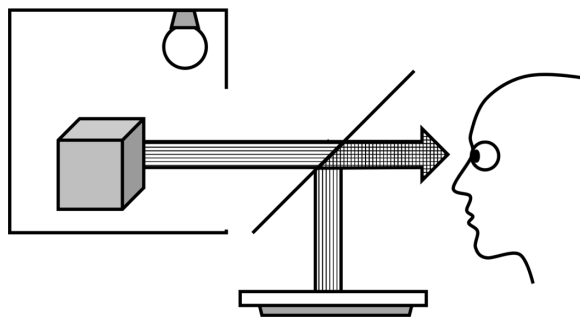


Fig. 1. AR setup. The observer, at left, sees an optical blend (indicated by the cross-hatched arrow) of real and virtual content in which the virtual content appears as a semitransparent, bright overlay. At the center of the diagram is a diagonal beam splitter that transmits light from the real-world objects illuminated in the viewing booth to the right (indicated by the horizontally hatched arrow) and reflects light from virtual content displayed on the LCD display below (indicated by the vertically hatched arrow).

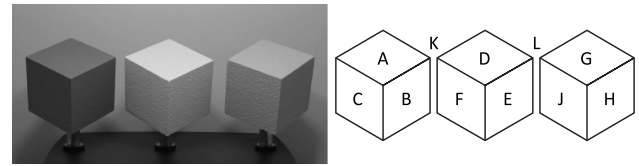


Fig. 2. At left, a photograph of the three cubes: dark, light, and mid gray. At right, the diagram indicates the labeled measurement positions of the three cubes' facets and viewing booth wall. These letter codes are used in Table 1.

The viewing booth is a metal box 85 cm wide, 50 cm tall, and 59 cm deep, with walls painted about 20% reflective gray. Illumination is provided by two Philips HUE A19 lamps, which are addressable RGB LEDs capable of a wide range of color and intensity. The LCD display is a Dell P2715Q 27-inch 4 K IPS LCD, addressed at 1920×1080 resolution by a PC running Windows 10 and MATLAB 2019a. A $45 \text{ cm} \times 33 \text{ cm}$ glass beam splitter is mounted at a 50 degree angle, with its center 56 cm above the LCD. From the perspective of the observer, the reflected LCD appears to float about halfway into the depth of the viewing booth, and it has a resolution of 55 pixels per degree of visual angle. The LCD was colorimetrically characterized for use in the experiment as described below.

B. Stimuli

The experimental stimuli included both real-world 3D cubes, illuminated with the LEDs described above, and virtual AR overlays, displayed with the LCD. The cubes were painted three shades of matte gray, with percent luminance factors (reflectances) of 13.1%, 50.5%, and 30.2%, respectively, at left, center, and right. Thus, the left cube can be referred to as dark gray, the center light gray, and the right mid gray. Because the illumination source was located at the top of the box, the luminances of the three visible facets of each cube were different. As measured with a Konica-Minolta CS-2000 spectroradiometer, from the point of view of the observer and through the beam splitter, the CIE 1931 XYZ tristimulus values (with Y in luminance, cd/m^2) of the nine visible cube faces and the rear wall of the viewing booth, with letter codes as in Fig. 2, are provided in Table 1. CIELAB coordinates were computed for each of the cubes' facets, using a hypothetical white reference (XYZ values 56.0, 57.6, and 45.4) equal to the measured XYZ of the center cube's top facet (D) divided by its measured reflectance. The top facets' L^* values were 43, 76, and 61, respectively, for left, center, and right.

AR overlays were created to fit the left and right cubes (dark and mid gray, respectively), adding light via the display, with the intention of raising their luminance to match or exceed that of the center light gray cube. Many levels of luminance for each of the overlays were precomputed, described further in the next section. From the perspective of the observer, the widths of the top facets of the cubes subtended 8.4, 8.3, and 8.1 degrees of visual angle, respectively, left to right. Six different types of AR overlays were generated, oversized by varying degrees from slight to obvious, as illustrated in Fig. 3:

A. Hex Faceted is a hexagon tightly fit to the cube with the three facets shaded to respectively maintain the physical

Table 1. Measured Absolute Tristimulus and CIE LAB Values of the Cube Facets and Viewing Booth Wall as Illuminated in the Viewing Booth without AR Overlay, with Y in Luminance, (cd/m^2)^a

Position	X	Y (cd/m^2)	Z	L^*	a^*	b^*
A	7.40	7.64	6.53	43.2	-0.29	-2.79
B	2.85	2.93	2.60	27.0	0.06	-3.00
C	2.31	2.38	2.01	24.11	-0.05	-1.61
D	28.2	29.1	22.9	76.4	0.0	0.0
E	7.06	7.25	5.79	42.1	0.21	-0.44
F	6.64	6.84	5.51	41.0	-0.04	-0.71
G	16.6	17.1	13.9	61.4	-0.12	-1.38
H	4.52	4.66	3.66	34.2	-0.07	0.10
J	6.51	6.70	5.61	40.2	-0.06	-1.91
K	12.5	12.8	9.01	54.3	0.26	4.67
L	12.8	13.1	9.19	54.8	0.30	4.67

^aBold rows A, D, and G correspond to the top facets of each cube used in matching; K and L correspond to the visually adjacent viewing booth wall.

luminance ratios of the original cube. This is the only overlay that is not uniform in color.

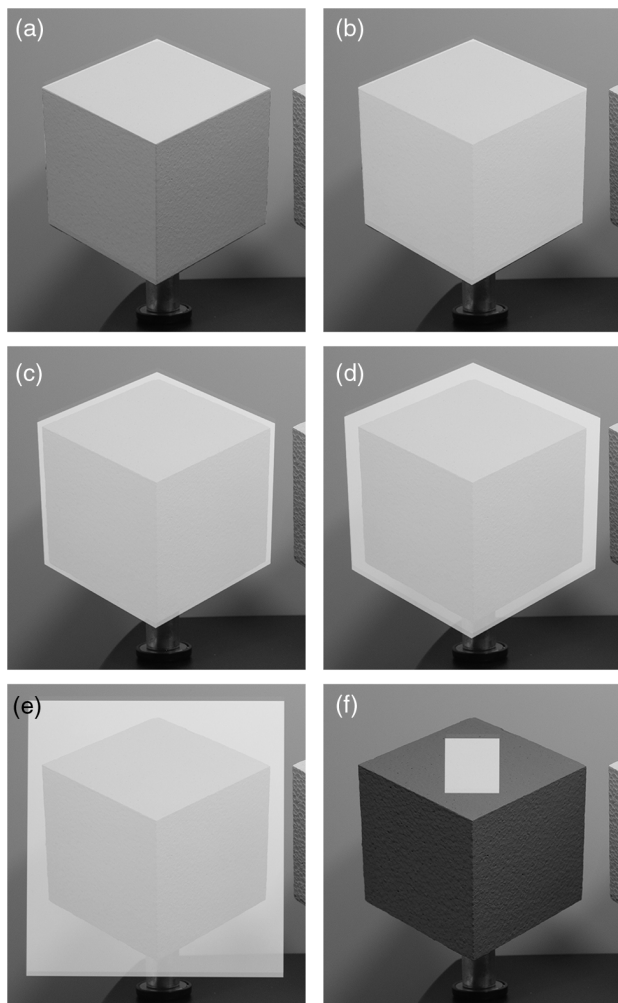


Fig. 3. Overlays. Photos of all six overlays shown on the left (dark gray) cube: labeled (a) Hex Faceted, (b) Hex Tight, (c) Hex Medium, (d) Hex Large, (e) Rect Large, and (f) Rect Tiny.

- B. Hex Tight is a uniform hexagon tightly fit to the cube. This differs from Hex Faceted because its facets are not shaded but uniform in color.
- C. Hex Medium is a uniform hexagon oversized larger than the cube by 0.17 degrees of visual angle on each side.
- D. Hex Large is a uniform hexagon oversized by 0.50 degrees of visual angle.
- E. Rect Large is a uniform rectangle of the same width as the Hex Large overlay.
- F. Rect Tiny is a uniform square of 2.0 degrees of visual angle, centered in the top facet of the cube.

The AR overlays were aligned to the real-world cubes by the author via a manual, keyboard-driven adjustment of the on-screen positions of the seven vertices of each cube (six perimeter corners plus the “center” nearest corner). The author performed this alignment visually using a dedicated graphical user interface at the beginning of the experiment and again after 10 observers because a slight misalignment was caused by a bump to the setup.

C. Observer Task

The experiment utilized the method of adjustment, with observers given a physical knob with which to adjust the intensity of the virtual AR overlay on one of the cubes. Specifically, observers were asked to add overlay intensity to match the brightness of the top face of one of the outer cubes (dark or mid gray) to the top face of the center cube (light gray); in each presentation there was only one overlay present as shown in the photograph in Fig. 4. Because the AR overlay is only capable of adding light, the overlays were used to make each of the outer darker cubes match the center light gray cube. For each overlay and each cube, levels of intensity were precomputed, chosen to create steps of 0.5 units of lightness L^* to cover a range from zero added light up to $L^* = 100$, where L^* corresponds to the proximal light stimulus, the sum of the darker cubes’ fixed, reflected luminance and the added luminance of the AR overlay itself. This range far exceeds the center cube’s physical reference L^* of 76.4, allowing observers headroom in their adjustment for the expected perceptual discounting. While not the same as brightness, steps of lightness were used to make the intensity levels much more perceptually uniform in spacing than they would have been in equal steps of luminance. This makes the method of adjustment easier and more efficient for the observer.

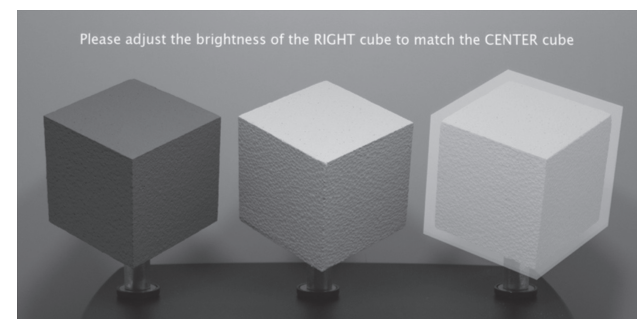


Fig. 4. Photograph of the observer’s field of view showing the instructions.

The intensity precomputation comprised steps of selecting a desired L^* value, computing the corresponding XYZ values, subtracting the measured XYZ values of each cube's top facet, and transforming the residual XYZ values to display-specific RGB drive values using an inverse display model. Due to slight spatial and angular inconsistencies, a different display characterization model was used for each cube, based on measurements made at the position of the center of each cube's top facet with the viewing booth LED illumination off. Using the method of Fairchild and Wyble, each model consisted of a black level offset, linear matrix, and lookup-table-based nonlinearity [20]. Model errors were quantified with measurement of 56 points near the neutral axis; the left and right models had mean errors in CIEDE2000 of 0.45 and 0.42 and max errors of 1.0 and 1.2, respectively, typical of contemporary LCDs.

Observers were instructed to use one eye to view the cubes through the fixed eyecup and were advised that if they switched eyes during the experiment, to pause for a minute or so to adapt. Before beginning, a short training session was provided, showing two different overlays on each side. During this, the task was explained: "Please adjust the brightness of the right/left cube to match the brightness of the center," after which the observer made and recorded their adjustment. The first training example was the Hex Faceted overlay, which is generally the easiest, as its shaded facets allow observers to match the whole cube. The second training was the Rect Large overlay, during which the experimenter clarified: "If it is impossible to make the entire cube match, as in this case, then please adjust the brightness of the top face of the right/left cube to match the brightness of the top face of the center cube." Subsequent training overlays were the Hex Large and Rect Tiny, and during the latter the further instruction was given: "In this case, your adjustment only affects the small rectangle; please adjust it to match the brightness of the top face of the center cube." Observers were asked if the task was clear and were given an opportunity to ask questions; all were satisfied and none had questions.

The experiment comprised 84 adjustments for each observer: two cubes (left and right) \times six overlays \times seven repetitions. Stimuli were presented in four blocks of 21 presentations, alternating left and right twice, and the starting side was counterbalanced over observers. The intention was for seven repetitions; however, due to the subdivision into four blocks, observers randomly saw six, seven, or eight repetitions of each condition, averaging seven. The time to complete the experiment ranged from 10 to 43 min with a median of 21 min. The experiment design and informed consent form were approved by RIT's Institutional Review Board.

D. Observers

Twenty-four observers voluntarily participated in the experiment, ranging in age from 19 to 64 with a median of 28. Gender was not recorded, nor color vision status, as they were expected not to impact the brightness-matching task with achromatic stimuli. Most of the observers were students, faculty, or staff of RIT. One observer's results were excluded because they were far outside the range of the other observers, implying the individual either did not understand the task or was using a unique matching criterion. The outlier observer's median response over

all experimental conditions was $L^* = 91.75$, while the other 23 observers' median responses ranged from 74.25 to 79.5; thus, the subsequent analysis was conducted with $N = 23$.

3. RESULTS

The method of adjustment task resulted in a single dependent variable, the observer-adjusted intensity (i.e., luminance) of the AR overlay that results in a brightness match with the center cube. The independent variables already mentioned were cube (left/dark gray and right/mid gray), and overlay (see Fig. 3). The AR overlays were presented on the LCD screen viewed through the beam splitter, with a measured relationship between display RGB and resulting luminance, and the resulting sum of the cubes' reflected luminance and the AR overlay luminance was transformed to CIELAB L^* (lightness). As explained above, lightness L^* was used, not brightness, because transformation to an estimate of brightness requires assumptions about adaptation and surround effects, while the CIELAB transformation simply requires the white point and is consistent for all cubes. Equal L^* values implies a physical match. Later in the analysis, appearance models were applied to account for differences between each cube's surroundings. An overview of the observer-adjusted L^* results is shown in Fig. 5, versus the six AR overlay types and split by the left/dark gray cube and right/mid gray cube. The split-violin plots show the distributions in L^* , slightly smoothed for clarity, of all observations for each overlay-cube combination ($N = \text{approx. } 161$ for each distribution). Plus signs indicate the mean values of each distribution with 95% confidence intervals, and the horizontal dashed line indicates the center reference cube L^* of 76.4: a physical match of the luminance of a darker cube plus AR overlay would lie on this line.

Several observations can be made. The leftmost overlays, Hex Faceted and Hex Tight, show the narrowest, most normal distributions, centered closest to the reference L^* , supporting the hypothesis that these tightly fit overlays would provide the most similar to physical matches. Said another way, the tightly fit overlays create no ambiguity or veiling glare effect and are presumably easiest for observers to match. From Hex Tight through Hex Medium, Hex Large, and Rect Large, there is an upward trend in mean L^* accompanied by a general widening of the distributions. This is stronger in the dark gray cube distributions than the mid gray cube distributions. This trend supports the hypothesis that an oversized overlay would lead the observer to add more light than physically necessary (thus higher L^* than the reference) in order to make a visual brightness match. Further, the gradual trend of higher L^* with overlay size shows no evidence for a discrete size threshold—such as the 0.25 degrees proposed by Fry and Alpern [6]—for the perception of veiling luminance. The Rect Tiny overlay is the only overlay for which the mean L^* of the dark gray cube is smaller than that of the mid gray cube. For this overlay only, the small rectangle being adjusted was immediately surrounded by the dark and mid gray cube tops, which would tend to make the apparent brightness higher and lower, respectively, due to simultaneous contrast.

Based on the distinct differences in the widths of the distributions in Fig. 5, both inter- and intraobserver standard deviations were computed as shown in Fig. 6. The interobserver

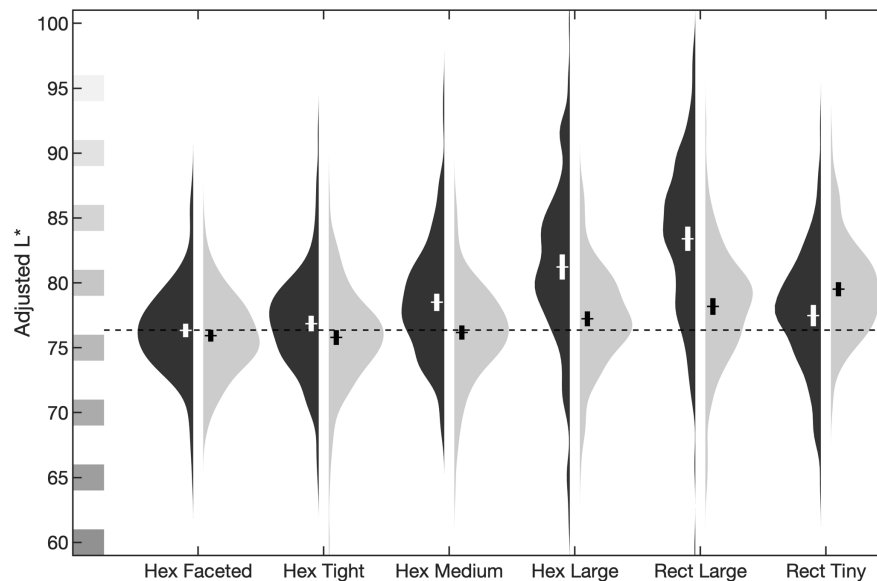


Fig. 5. Overall results shown in a split-violin plot. Smoothed distributions of observer-adjusted CIELAB L^* are plotted for each of the six named AR overlay types on the x axis. The darker left distributions show the results for the left/dark gray cube, and the lighter right distributions show the results for the right/mid gray cube. Contrasting + indicate mean values of each distribution, also indicating 95% confidence intervals with their height. The dashed horizontal line shows the measured L^* of 76.4 of the top facet of the center cube, the reference for all visual matches.

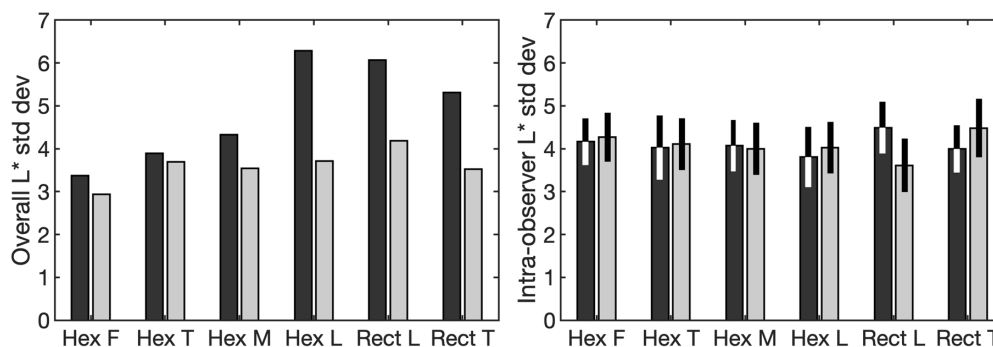


Fig. 6. Standard deviation of adjusted L^* values: (left) overall standard deviations for each overlay-cube combination and (right) intraobserver standard deviations: the means of standard deviations computed for each observer, with 95% confidence intervals. Darker bars refer to the left dark gray cube, and lighter bars to the right mid gray cube.

standard deviations in the left plot reflect the distribution widths apparent in Fig. 5, while the intraobserver standard deviations in the right plot remain fairly consistent both in height and in spread; this suggests that observers may employ divergent matching criteria in the larger overlay cases. The experiment procedure did not include formal follow-up questions, but informal discussions with the observers indicated that they generally found the larger overlays more difficult to match. This could explain the widening of corresponding distributions in the matching results.

A. ANOVA

Several analyses of variance (ANOVAs) were computed using MATLAB. In what will be referred to as the overall ANOVA, observers' adjusted L^* was the dependent variable, and cube, overlay, and their interaction were specified as fixed factors.

Additionally, observer was specified as a random factor to estimate the variation attributable to different individuals. In the overall ANOVA, all factors were all found to have significant effects with $\alpha = 0.05$, as shown in Table 2, justifying the rejection of the null hypothesis H_0 . Subsequently, the data were split by cube, and two separate ANOVAs were run to elucidate the differences between overlays. Both overlay and observer were found significant for both cubes as seen in Tables 3 and 4.

Tukey HSD post-hoc tests were conducted between overlay types for each cube's ANOVA results, which clarify the differences in L^* seen in Fig. 5: for the left dark gray cube, Rect Large and Hex Large were each found to be significantly different than all other overlays, while Hex Medium was found to be significantly larger than both Hex Faceted and Hex Tight. Hex Faceted, Hex Tight, and Rect Tiny were found to not significantly differ from one another. For the right mid gray cube, Rect Tiny was found to be significantly larger than all other overlays, while Rect Large was found to be significantly larger than all of

Table 2. Overall ANOVA Results for the Entire Dataset^a

Source	Sum Sq.	d.f.	Mean Sq.	<i>F</i>	<i>p</i>
Cube	1623.4	1	1623.35	97.19	<0.001
Overlay	5231.2	5	1046.24	62.64	<0.001
Observer	4746.4	22	215.74	12.92	<0.001
Cube* overlay	2810	5	562	33.65	<0.001
Error	31,703.1	1898	16.7		
Total	46,006	1931			

^a H_0 is rejected for all factors.**Table 3. ANOVA Results for the Left, Dark Gray Cube^a**

Source	Sum Sq.	d.f.	Mean Sq.	<i>F</i>	<i>p</i>
Overlay	6191.2	5	1238.25	62.39	<0.001
Observer	5250.8	22	238.67	12.03	<0.001
Error	18,616.1	938	19.85		
Total	30,077.9	965			

^a H_0 is rejected for both factors.**Table 4. ANOVA Results for the Right, Mid Gray Cube^a**

Source	Sum Sq.	d.f.	Mean Sq.	<i>F</i>	<i>p</i>
Overlay	1781.7	5	356.347	34.66	<0.001
Observer	2938.7	22	133.577	12.99	<0.001
Error	9643.8	938	10.281		
Total	14,338.4	965			

^a H_0 is rejected for both factors.

the smallest three (Hex Tight, Hex Faceted, and Hex Medium). Hex Large was found to be significantly larger than Hex Tight and Hex Faceted.

In order to identify significant differences between the cubes' adjusted L^* for each overlay, another ANOVA was computed, similar to the overall ANOVA except that cube \times overlay was coded as a single factor. This was followed by a Tukey HSD post-hoc test on the cube \times overlay factor to assess pairwise comparisons. Left dark and right mid cubes were found to be significantly different from one another with overlays Hex Medium, Hex Large, Rect Large, and Rect Tiny. No significant difference was found for Hex Faceted and Hex Tight.

B. Interobserver Differences

The ANOVA results above consistently show significant effects between observers, so several analyses were performed to look for reasons for this; however, no attributable effect was found, meaning that there are simply differences between observers in their task performance. First, the effect of observer age was tested. The overall ANOVA on the L^* results with fixed factors overlay, cube, and their interaction was extended to include observer age nested within observer, with age and observer both designated random factors. All factors were significant ($p < 0.001$) except age ($p = 0.61$), indicating that the null

hypothesis—that age is not related to the differences—cannot be rejected.

A learning effect was presumed in the pilot study mentioned previously, but it was confounded with cube presentation order [19]. With this in mind, the present experiment was designed to subdivide the presentations into four blocks, alternating left and right twice, and to counterbalance the starting side over observers. Comparing the first half and second half of each observer's results, an ANOVA was run on the L^* results with fixed factors overlay, cube, their interaction, half, and observer as a random factor. All factors were significant ($p < 0.001$) except half ($p = 0.086$), which implies that though there is a small mean difference between L^* in the first half and second half (77.9 versus 78.2), the null hypothesis cannot be rejected. This implies that there is no evidence for a significant learning effect. Further, when the data were split by half and analyzed as in the overall ANOVA, the results of both halves showed the same significant effects mentioned previously. This reinforces that any learning effect present is trivial, as the same conclusions would be drawn by looking only at the data from either half.

Looking at presentation order, the overall ANOVA on the L^* results with fixed factors overlay, cube, and their interaction was extended to include starting cube nested within observer, with starting cube and observer both designated random factors. All factors were significant ($p < 0.001$) except starting cube ($p = 0.77$), indicating that starting cube is not related to the differences.

Finally, because the randomization in the experimental interface did not consistently present seven repetitions of every overlay-cube combination to each observer, the data were trimmed to include only the first six repetitions, resulting in a slightly smaller dataset. The results were only slightly different than those from the full dataset, with mean L^* per overlay-cube combination moving on average 0.10 units. An overall ANOVA on the trimmed dataset revealed the same significant effects as the full dataset, indicating that including or not including the varying number of repetitions does not matter.

In summary, statistically significant differences in brightness matches were found between the AR overlays, between the dark gray and mid gray cubes, and between observers. Further analysis clarified that the AR overlay differences are stronger for the dark gray cube than the mid gray cube. There are no significant differences due to observer age, learning effects, nor presentation order.

4. MODELING AND DISCUSSION

A. Appearance Modeling

With results and effects clear, the question becomes whether conventional appearance predictions can explain them and/or whether there are effects specific to the OST-AR presentation. CIECAM02 is an appearance model that takes into account surround luminance as well as absolute adapting luminance to predict appearance correlates including lightness (J), brightness (Q), chroma (C), and hue (b) [5]. CIECAM02 J and Q values were computed from the adjusted L^* values discussed previously, first reverting to physical XYZ values, then using the white XYZ values 56.0, 57.6, and 45.4. The CIECAM02 input parameters were surround “average,” adapting luminance

$L_A = 13.0 \text{ cd/m}^2$ (average of measured viewing booth wall luminance values K and L in Table 1), and relative background luminance Y_b chosen depending on the overlay. CIECAM02 uses the parameter Y_b to describe the background luminance (here “background” refers to the surrounding region, different from the use of “background” elsewhere in this paper, which describes the real-world scene transmitted through the AR system) in the region up to 10 degrees of visual angle beyond the 2 degree stimulus. In this experiment, the cube stimuli were approximately 8 degrees in size, and the overlays extended 0–0.5 degrees beyond the cubes. For the tight overlays with zero extension (Hex Faceted and Hex Tight), Y_b was fixed at the viewing booth wall luminance L_A . For the Rect Tiny overlay, Y_b was fixed at the luminance of the top of the respective dark and mid gray cube. For the others, with oversized overlays, Y_b depended on the adjusted overlay intensity and was the viewing booth wall + overlay luminance. CIECAM02 values were computed with these divergent Y_b values, a best estimate of the physical environment expected to affect appearance.

The objective of this modeling was to determine if CIECAM02 provides an appropriate prediction of the measured appearance matches, which would result in values equal to the reference in all cases. In this analysis, each individual observation was converted to CIELAB and CIECAM02 values, but only mean values with 95% confidence intervals are compared in the plots. Figure 7 compares luminance (Y), CIELAB lightness (L^*), CIECAM02 lightness (J), and CIECAM02 brightness (Q). For clarity, differences between the mean results in each unit and their respective reference values are shown; also, because each unit has a different absolute magnitude, the y axes are scaled relative to the respective reference value. If any of these units were accurate predictors of the observed matches, the plotted differences would have zero magnitude. However, in all four plots the results still differ between overlay-cube conditions,

which says that the J and Q computations of CIECAM02, which aim to account for the surround effects, do not sufficiently explain the intercondition deviations. Surprisingly, the Q bars [plot (d)] are the largest, with highly negative values implying that CIECAM02 underestimates the brightness of the cubes being adjusted. Perhaps the small size of the bright surrounds provided by the oversized overlays in the Hex M, Hex L, and Rect L conditions does not depress perceived brightness as much as the larger surrounds for which CIECAM02 was optimized. The L^* and J plots [plots (b) and (c)] are similar in overall magnitude, though notably J indicates a clear reduction in the difference between the cubes with the Rect Tiny overlay [the height difference between the pairs of Rect T bars in plot (c), compared with that in plot (b)], which is a simultaneous contrast-like arrangement that is most similar to the 2 degree stimulus and 10 deg background that CIECAM02 describes. However, the J values are not closer to the reference.

Because the effect of the oversized overlays appears to strongly influence the CIECAM02 results, a small test was made with the Hex Medium condition. As described above, its relative background luminance Y_b was computed from the overlay as with the larger overlays; however, the small size of the overlay means it could have been treated like the smaller overlays, using the surround luminance. Both computations were tried—the plot shows the former, erring below the reference, while the latter pushes the error above the reference; unfortunately, it is not clear that one is superior to the other, and maybe an intermediate value would be more appropriate. In any case, neither CIECAM02 lightness J nor brightness Q predictions resolve the increasing overlay intensity with larger overlays, nor the differences between cube reflectance factor. Accounting for surround effects with a conventional appearance model is not sufficient.

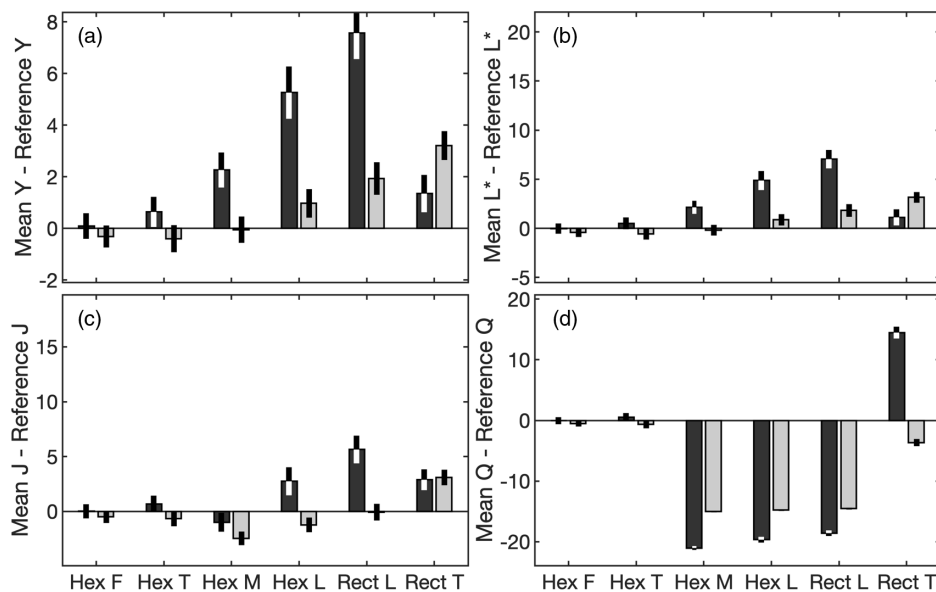


Fig. 7. Differences between observed mean and reference values expressed in (a) luminance Y , (b) CIELAB lightness L^* , (c) CIECAM02 lightness J , and (d) CIECAM02 brightness Q , with 95% confidence intervals. Each plot shows pairs of bars for each overlay; darker bars refer to the left dark gray cube, and lighter bars to the right mid gray cube. For fair comparisons, vertical axes are scaled relative to magnitude of each reference value, and plot (d) is shifted upward to make room for the negative differences.

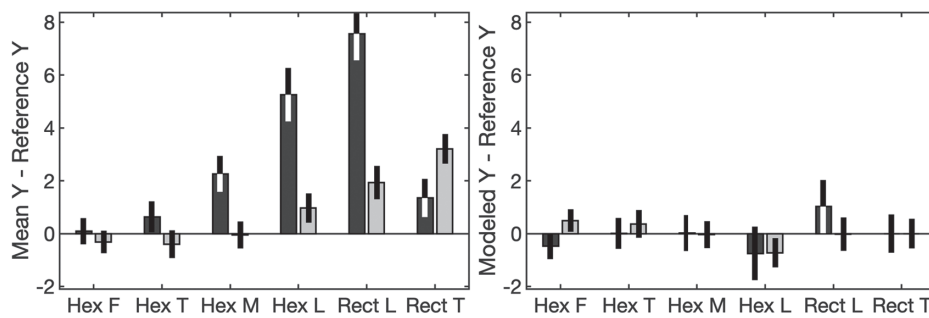


Fig. 8. Differences between observed mean and reference values for (left) physical luminance values and (right) modeled effective luminance per overlay-cube combination, with 95% confidence intervals. Darker bars refer to the left dark gray cube, and lighter bars to the right mid gray cube.

B. Foreground-Background Discounting

Previous studies also found conventional appearance models insufficient to describe observers' matches in AR; thus, Hassani's proposed model of a non-physical weighted sum of AR foreground and real-world scene background may be tested [4]. The structure of her formulation was expressed in Eq. (2). Hassani's research found β smaller than α because observers discounted the background contribution, but the opposite is expected in the present study because observers presumably discounted the foreground to match the background objects. The present results show a clear effect of overlay size on the amount of extra light added to create a brightness match. This means the degree of discounting changes with overlay, and therefore α must depend on the extent of the oversized overlay: one α for the two overlays Hex Faceted and Hex Tight, another α for Hex Medium, and a third α for Hex Large and Rect Large. A common β value can account for the background cube reflectance difference. Rect Tiny behaves differently (more light for the mid gray cube than the dark gray cube), so it was left out and given its own α and β , which of course guarantees a perfect fit.

To test this approach, α and β values for Eq. (2) were fit using a matrix inversion, setting XYZ_{eff} to the measured reference cube data, setting XYZ_{BG} to the mid gray and dark gray cube data, respectively, and computing XYZ_{FG} from the mean observer matches. The result is a very good prediction of the discounting effect, although the number of data points is not much greater than the number of degrees of freedom in the fit. Accounting for the cube differences, the fitted β value is 1.09. For the AR overlays, fitted α values decrease with increasing size: 0.95, 0.88, and 0.76, meaning further discounting as the overlay is more oversized, as expected, and confirming that $\alpha < \beta$ as expected. The Hex Fitted and Hex Tight overlays, with (α, β) of (0.95, 1.09), are the closest to a physical match (1, 1), which implies that observers could indeed (nearly) interpret the overlays as manipulating the object brightness. The Rect Tiny (α, β) are (1.01, 0.80), with $\alpha > \beta$, indicating that the background is discounted, similar to Hassani's results with small foreground patches on larger backgrounds.

The left panel of Fig. 8 shows the measured mean luminance values [corresponding to the mean L^* values plotted in Fig. 5 and repeating plot (a) in Fig. 7] for all 12 overlay-cube combinations, nearly all of which are higher than the reference luminance. The right panel shows the modeled effective luminance values, computed with the discounting model,

which are all very close to the reference. Though plotted in absolute luminance, the differences from the reference can correspondingly be summarized in $|\Delta L^*|$: the mean for the measured values is 2.15, and for the modeled values it is 0.35. Thus, the present study provides further support for the concept of foreground-background discounting as modeled with the structure of Eq. (2).

C. Future Work

The present experiment provides initial data on the effect of AR overlay size and alignment on the perceived lightness of real-world objects, supporting the AR foreground-background discounting model. However, the goal remains to generalize this model beyond one-dimensional lightness and for a wider range of situations. More visual matching data for more foreground-background color combinations will be needed to exercise and verify the model structure. An important focus would be to deterministically choose the α and β weighting scalars for various situations. Additionally, less constrained viewing environments should be tested, including stereo presentations and presentations that allow head motion with properly rendered parallax. Temporal variations in the alignment of AR overlays may lead to further discounting.

5. CONCLUSION

The results of this experiment quantify the appearance effects seen when beam-splitter-generated OST-AR overlays, transparent and additive in nature, are used to manipulate the perceived brightness of real-world objects. The size and alignment of the overlays, as well as the physical reflectance of real-world cubes, were shown to affect the amount of extra light added to visually match the brightness of a lighter reference cube. The extra light implies a discounting mechanism by which the observers interpret the AR overlay as a veiling luminance and partially ignore it, supporting the general use of a foreground-background discounting model describing a visual equivalent but non-physical weighting of foreground and background contributions. AR foreground discounting is minimized with tightly fit, physically matching AR overlays. Foreground discounting increases with increasingly oversized AR overlays, and it inverts (meaning it becomes background discounting as in previous work) with undersized AR overlays. The amount of discounting increases gradually with size in the range tested, rather than showing

any discrete threshold at which veiling luminance is perceived. These results will help design AR systems and interfaces, providing insight into tolerances for fitting virtual content to real-world scenes and predicting how users will interpret visual mixtures of real and virtual stimuli.

Funding. National Science Foundation (1942755).

Acknowledgment. The author thanks RIT alumna Sara Leary for her excellent work on the pilot study, including coding an early version of the matching experiment interface, constructing and painting the cubes, and preparing the SAP poster presentation [19].

Disclosures. The author declares no conflicts of interest.

REFERENCES

1. J. L. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester, "More than meets the eye: an engineering study to empirically examine the blending of real and virtual color spaces," in *IEEE Virtual Reality Conference (VR)* (2010), pp. 79–86.
2. C. Menk and R. Koch, "Truthful color reproduction in spatial augmented reality applications," *IEEE Trans. Visual. Comput. Graph.* **19**, 236–248 (2013).
3. N. Hassani and M. J. Murdoch, "Investigating color appearance in optical see-through augmented reality," *Color Res. Appl.* **44**, 492–507 (2019).
4. N. Hassani, "Modeling color appearance in augmented reality," Ph.D. dissertation (Rochester Institute of Technology, 2019).
5. "A Colour Appearance Model For Colour Management Systems: CIECAM02," CIE 159:2004 (Commission Internationale de L'Eclairage, 2004).
6. G. A. Fry and M. Alpern, "The effect of veiling luminance upon the apparent brightness of an object," *Optometry Vision Sci.* **31**, 506–520 (1954).
7. A. L. Gilchrist and A. Jacobsen, "Lightness constancy through a veiling luminance," *J. Exp. Psychol.* **9**, 936–944 (1983).
8. A. Soranzo, A. Galmonte, and T. Agostini, "The luminance misattribution in lightness perception," *Psihologija* **43**, 33–45 (2010).
9. F. A. A. Kingdom, "Perceiving light versus material," *Vision Res.* **48**, 2090–2105 (2008).
10. F. A. A. Kingdom, "Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy," *Vision Res.* **51**, 652–673 (2011).
11. V. Ekroll, F. Faul, and R. Niederée, "The peculiar nature of simultaneous colour contrast in uniform surrounds," *Vision Res.* **44**, 1765–1786 (2004).
12. V. Ekroll and F. Faul, "Transparency perception: the key to understanding simultaneous color contrast," *J. Opt. Soc. Am. A* **30**, 342–352 (2013).
13. F. Metelli, "The perception of transparency," *Sci. Am.* **230**, 90–99 (1974).
14. M. D'Zmura, P. Colantoni, K. Knoblauch, and B. Laget, "Color transparency," *Perception* **26**, 471–492 (1997).
15. D. Zavagno, O. Daneyko, and K. Sakurai, "What can pictorial artifacts teach us about light and lightness?" *Jpn. Psychol. Res.* **53**, 448–462 (2011).
16. R. S. Berns, *Billmeyer and Saltzman's Principles of Color Technology*, 4th ed. (Wiley, 2019).
17. CIE DIS 017/E:2016 ILV: International Lighting Vocabulary, 2nd ed. (Commission Internationale de L'Eclairage, 2016), <http://eilv.cie.co.at/>.
18. A. Gilchrist, "Theoretical approaches to lightness and perception," *Perception* **44**, 339–358 (2015).
19. S. Leary and M. J. Murdoch, "Manipulating object lightness in augmented reality," in *ACM Symposium on Applied Perception* (ACM, 2018).
20. M. Fairchild and D. Wyble, "Colorimetric characterization of the Apple studio display (Flat panel LCD)," Munsell Color Science Laboratory Technical Report (1998), <https://scholarworks.rit.edu/article/920/>.